Scalings of Matrices Satisfying Line-Product Constraints and Generalizations

Uriel G. Rothblum
Faculty of Industrial Engineering and Management
Technion—Israel Institute of Technology
Haifa 32000, Israel
and
RUTCOR—Rutgers Center for Operations Research
Rutgers University
P.O.B. 5062
New Brunswick, New Jersey 08904

and

Stavros A. Zenios
Decision Sciences Department
The Wharton School
University of Pennsylvania
Philadelphia, Pennsylvania 19104

Submitted by Hans Schneider

ABSTRACT

A *scaling* of a nonnegative matrix A is a matrix having the form $A' = UAV$ where U and V are square diagonal matrices which have positive diagonal elements. If the matrix A is square and $V = U^{-1}$, we call A' *a symmetric scaling* of A. We consider two problems: the first concerns the identification of a scaling of a given nonnegative matrix with prescribed row and column products; the second concerns the problem of finding a symmetric scaling of a given nonnegative square matrix whose row products equal the corresponding column products. For each of these two scaling problems we characterize the solutions in terms of a nonlinear convex optimization problem, we use the characterization to demonstrate that feasibility of either scaling problem is equivalent to the existence of a matrix satisfying the target property and having the same pattern as the given matrix, we establish uniqueness of solutions to either scaling problem whenever it is feasible, and we develop algorithms for computing the desired (unique) scalings in the cases where the problems are feasible.

655 Avenue of the Americas, New York, NY 10010 0024-3795/92/$5.00

1. INTRODUCTION

A *scaling* of a nonnegative matrix A is a matrix having the form $A' = UAV$ where U and V are square diagonal matrices which have positive diagonal elements. If the matrix A is square and $V = U^{-1}$, we call A' a *symmetric scaling* of A. Scaling problems concern the identification of scalings of a given matrix with certain prescribed properties. In the current paper we consider scaling problems where the objective is related to row and column products. Specifically, we study two problems: the first concerns the identification of a scaling of a given matrix with prescribed row and column products, whereas the second concerns the identification of a symmetric scaling whose row products equal the corresponding column products.

The problem of determining a scaling of a given matrix that has prescribed row and column sums has been studied extensively over the last half century. The first reference of which we are aware on such scaling problems is Kruithof (1937), who considered estimation methods for telephone traffic. In particular, he suggested an algorithm for computing such a scaling for a given nonnegative matrix by scaling the rows and the columns iteratively to the right sums. The development of a comprehensive theory started only through the pioneering paper of Sinkhorn (1964). Major advances followed, e.g., Brualdi, Parter, and Schneider (1966), Knopp (1979), Brualdi (1968), Menon (1968), Menon and Schneider (1969), and others. The problem has applications in many areas, including planning of telephone and transportation traffic, updating social accounting matrices, matrix preconditioning, and algebraic image reconstruction; see Bacharach (1970), Lamond and Stewart (1981), King (1981), Rothblum and Schneider (1989), Schneider and Zenios (1990), and references therein. Recently, Rothblum, Schneider, and Schneider (1990) considered the related problem of identifying a scaling of a given matrix with prescribed row and column maxima. In particular, they obtained an algorithm which, for a given matrix, will either identify a desired scaling or determine that no such scaling exists. In the current paper we consider the related scaling problem where the objective is to identify a scaling of a given matrix having prescribed row and column products.

Symmetric scalings of a given square matrix whose row sums equal the corresponding column sums have been studied in Eaves, Hoffman, Rothblum, and Schneider (1985), and a comprehensive study of computational methods for the numerical solution of such problems is given in Schneider and Zenios (1990). The related problem with the max operator replacing the sum operator, i.e., the problem of finding symmetric scalings of a given square matrix whose row maxima equal the corresponding column maxima, was introduced in Schneider and Schneider (1991) and further studied in

Schneider and Schneider (1990), Schneider (1990), and Rothblum, Schneider, and Schneider (1992). Here, we consider the related problem of identifying symmetric scalings of a given square matrix whose row products equal the corresponding column products.

Scaling problems where the target property is that the row sums and column sums satisfy prescribed lower and upper bounds have been examined in Rothblum (1990). Numerical algorithms for solving these problems which rely on converting them to equivalent optimization problems are given in Censor and Zenios (1991). The above two references also address the symmetric scaling problem where the target property is that the differences between row sums and the corresponding column sums have prescribed lower and upper bounds.

The analysis of the scaling problems we study in the current paper follows the same outline as Eaves, Hoffman, Rothblum, and Schneider (1985), Rothblum and Schneider (1989), and Rothblum (1989, 1992), where scaling problems with constraints on row and column sums were studied. In the above papers nonlinear convex optimization problems were used to obtain results about characterization, existence, and uniqueness of the scaling problems. Here, we identify a least squares problem under a set of linear equality constraints whose solution is equivalent to the solution of the scaling problem with constraints on row and column products. In particular, results about characterization, existence, and uniqueness of solutions to scaling problems are derived from corresponding results about solutions of least squares problems.

We start by introducing some notation and conventions in Section 2. Next, in Section 3, we introduce and analyze a model which is later used to unify the study of the two scaling problems we consider in this paper—finding scalings of given matrices with prescribed row and column products, and finding symmetric scalings of given square matrices with row products equaling the corresponding column products. These two problems are then examined in Sections 4 and 5, respectively. Finally, explicit algorithms for solving the above problems are suggested in Section 6.

2. NOTATION AND CONVENTIONS

A vector $a \in R^p$ is called *nonnegative*, written $a \geqslant 0$, if all of its coordinates are nonnegative, and *strictly positive*, written $a \gg 0$, if all of its coordinates are positive. Given two vectors $a, b \in R^n$, we write $a \geqslant b$ or $a \gg b$ if $a - b \geqslant 0$ or $a - b \gg 0$, respectively. Similar notation applies to matrices.

Let $A \in R^{m \times n}$. For $i = 1, \ldots, m$, the *i-row-support* of A, denoted $\sigma_i(A)$, is defined as the set $\{j = 1, \ldots, n : A_{ij} \neq 0\}$; similarly, for $j = 1, \ldots, n$, the *j-column-support* of A, denoted $\sigma^j(A)$, is defined as the set $\{i = 1, \ldots, m : A_{ij} \neq 0\}$. The *support* of A, denoted $\sigma(A)$, is defined as the set $\bigcup_{i=1}^m \{(i,j) : j \in \sigma_i(A)\} = \bigcup_{j=1}^n \{(i,j) : i \in \sigma^j(A)\}$. If I is a subset of $\{1, \ldots, m\}$ and J is a subset of $\{1, \ldots, n\}$, we denote the *submatrix of A corresponding to* the rows of I and the columns of J by A_{IJ}; similarly, we denote the submatrix of A corresponding to the rows not in I and the columns not in J by $A_{I^c J^c}$.

The *bipartite graph associated with* the matrix A is the pair $G = (N, E)$ where $N \equiv \{1, \ldots, m + n\}$ is the *vertex set* of G and

$$E \equiv \big\{\{i, m + j\} : (i, j) \in \sigma(A)\big\}$$

is the *edge set* of G. We can order the edges of E, say lexicographically, and identify each edge with a corresponding (unique) integer in the set $\{1, \ldots, |E|\}$. The *node-arc incidence matrix* of G is then defined as the matrix $C \in R^{|N| \times |E|}$ with

$$C_{ie} = \begin{cases} 1 & \text{if} \quad e = \{i, j\} \text{ for some } j = 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, if $u \in R^m$, $v \in R^n$, and $e = \{i, m + j\} \in E$, then $[(u^T, v^T)C]_e = u_i + v_j$. Also, if $a \in R^{|E|}$, then for every vertex $k \in V$, $(Ca)_k = \sum_{\{e : k \in e\}} a_e$.

The *directed graph associated with* a (square) matrix $A \in R^{n \times n}$ is the pair $H = (N, E)$ where $N \equiv \{1, \ldots, n\}$ is the *vertex set* of H and

$$E \equiv \big\{(i, j) : i = 1, \ldots, n \quad \text{and} \quad j = 1, \ldots, n, \text{where } A_{ij} \neq 0\big\}$$

is the *edge set* of H. Again, we can order the edges of E, say lexicographically, and identify each edge with a corresponding (unique) integer in the set $\{1, \ldots, |E|\}$. The *node-arc incidence matrix* of H is then defined as the matrix $C \in R^{|N| \times |E|}$ with

$$C_{ie} = \begin{cases} 1 & \text{if} \quad e = (i, j) \text{ for some } j \in \{1, \ldots, n\}, \\ -1 & \text{if} \quad e = (j, i) \text{ for some } j \in \{1, \ldots, n\}, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, if $w \in R^n$ and $e = (i, j) \in E$, then $(w^T C)_e = w_i - w_j$. Also, if $a \in R^{|E|}$, then for every vertex $i \in N$, $(Ca)_i = \sum_{\{j : (i, j) \in E\}} a_{ij}$.

3. A UNIFIED MODEL

Throughout this section let $a \in R^p$, $b \in R^q$, and $C \in R^{p \times q}$ be given. We consider the problem of finding vectors $a' \in R^p$ and $w \in R^q$ such that

$$a'^T = a^T + w^T C \tag{3.1}$$

and

$$Ca' = b. \tag{3.2}$$

Thus, the goal is to select a vector $w \in R^q$ so that the transformation $a \to a + C^T w$ maps the vector a into a vector a' that satisfies the target system $Cx = b$. The vectors a and b and the matrix C form the data for this problem. In particular, the matrix C serves two roles—it is used both in the definition of the transformation and in the target condition.

We will see that the problem of finding solutions to (3.1)–(3.2) is facilitated by examining the following (nonlinear, convex) least squares optimization problem:

PROGRAM I. $\min 2^{-1}\sum_{j=1}^{p}(x_j - a_j)^2$

$$\text{subject to} \qquad Cx = b. \tag{3.3}$$

The relation of the above program to (3.1)–(3.2) will be explored in the following three theorems.

THEOREM 3.1 (Existence). *Let $a' \in R^p$. Then there exists a vector $w \in R^q$ such that (a', w) satisfies* (3.1)–(3.2) *if and only if a' is an optimal solution of Program* I.

Proof. As the objective of Program I is strictly convex and its constraints are defined via linear equalities, we have that the Kuhn-Tucker conditions are necessary and sufficient for optimality; see Avriel (1976). For Program II, the Kuhn-Tucker conditions assert that $x \in R^p$ satisfies

$$Cx = b$$

and for some vector $w \in R^q$

$$0 = \frac{\partial}{\partial x_j}\left[2^{-1}\sum_{j=1}^{p}(x_j - a_j)^2 - w^T(Cx - b)\right] = x_j - a_j - (w^T C)_j$$

$$\text{for} \quad j = 1, \dots, n.$$

Thus, the vector a' is an optimal solution of Program I if and only if for some vector $w \in R^q$, (3.1)–(3.2) is satisfied by (a', w). ■

THEOREM 3.2 (Existence). *The following are equivalent*:

(a) *there exists a pair* $(a', w) \in R^q \times R^p$ *satisfying* (3.1)–(3.2),
(b) *the linear system* $Cx = b$ *is feasible*,
(c) *Program* I *has an optimal solution*, *and*
(d) *there exists no vector* $\lambda \in R^q$ *with* $\lambda^T C = 0$ *and* $\lambda^T b \neq 0$.

Proof. The equivalence of (a) and (c) follows directly from Theorem 3.1. Also, the implication (c) $\Rightarrow$ (b) is trivial, and the reverse implication (b) $\Rightarrow$ (c) follows from continuity arguments and the observation that the objective function of Program I has compact level sets. Finally, the equivalence of (b) and (d) is the standard result in linear algebra which asserts that for every matrix C, range(C) = null$(C^T)^{\perp}$. ■

THEOREM 3.3 (Uniqueness). *There exists at most one vector* $a' \in R^p$ *for which there exists a vector* $w \in R^q$ *such that* (a', w) *satisfies* (3.1)–(3.2). *Further*, *if* $(a', w^1) \in R^p \times R^q$ *satisfies* (3.1)–(3.2), *then the general solution of* (3.1)–(3.2) *is* $(a', w^1 + z)$, *where* $z \in R^q$ *is a solution of the homogeneous equation* $z^T C = 0$.

Proof. The uniqueness of the vector a' follows directly from Theorem 3.1 and the fact that the objective function of Program I is strictly convex. Next assume that (3.1)–(3.2) is satisfied by $(a', w^1) \in R^p \times R^q$. Then $(a', w^2) \in R^p \times R^q$ satisfies (3.1)–(3.2) if and only if $a + (w^2)^T C = a'$. As $a + (w^1)^T C = a'$, it follows that the latter occurs if and only if $(w^2 - w^1)^T C = a' - a' = 0$, i.e., $w^2 = w^1 + z$ where $z^T C = 0$. ■

The next lemma complements Theorem 3.1 by characterizing vectors w, rather than a', which can be extended into a solution of (3.1)–(3.2). This result is not used in our forthcoming analysis of scaling problems and is stated for completeness.

LEMMA 3.4. *Let* $w \in R^q$. *Then there exists a vector* $a' \in R^p$ *such that* (a', w) *satisfies* (3.1)–(3.2) *if and only if* $CC^T w = b - Ca$.

Proof. The proof of the lemma follows directly from the substitution of (3.1) into (3.2). ■

4. SCALINGS OF MATRICES HAVING PRESCRIBED ROW AND COLUMN PRODUCTS

Throughout this section let $A \in R^{m\times n}$ be a given nonnegative matrix having no zero row or zero column, and let $S \in R^m$ and $T \in R^n$ be given strictly positive vectors. We consider the problem of finding a matrix $A' \in R^{m\times n}$ and diagonal matrices $U \in R^{m\times m}$ and $V \in R^{n\times n}$ with positive diagonal elements such that

$$A' \equiv UAV \in R^{m\times n}, \tag{4.1}$$

$$\prod_{j\in\sigma_i(A)} A'_{ij} = S_i \qquad \text{for} \quad i = 1,\dots,m, \tag{4.2}$$

and

$$\prod_{i\in\sigma^j(A)} A'_{ij} = T_j \qquad \text{for} \quad j = 1,\dots,n. \tag{4.3}$$

We note that by defining the empty product to be zero, we may allow the matrix A to have zero rows and/or zero columns and the vectors S and T to be nonnegative, rather than positive. But, under this generalization, solvability of (4.1)–(4.3) implies that all the elements of a row or a column of A are zero if and only if the corresponding coordinate of S or T is zero, and in this case we can omit the zero rows and columns of A and the corresponding coordinates of S and T, respectively.

Consider the matrix $a \in R^{m\times n}$ and the vectors $s \in R^m$ and $t \in R^m$ defined elementwise by

$$a_{ij} \equiv \begin{cases} \ln A_{ij} & \text{if} \quad (i,j) \in \sigma(A), \\ 0 & \text{if} \quad (i,j) \in (\{1,\dots,m\} \times \{1,\dots,n\}) \setminus \sigma(A), \end{cases} \tag{4.4}$$

$$s_i \equiv \ln S_i \qquad \text{for} \quad i = 1,\dots,m, \tag{4.5}$$

and

$$t_j \equiv \ln T_j \qquad \text{for} \quad j = 1,\dots,n. \tag{4.6}$$

Then (4.1)–(4.3) is equivalent to the problem of finding a matrix $a' \in R$ and vectors $u \in R^m$ and $v \in R^n$ such that

$$a'_{ij} = \begin{cases} u_i + a_{ij} + v_j & \text{if } (i,j) \in \sigma(A), \\ 0 & \text{if } (i,j) \in (\{1,\dots,m\} \times \{1,\dots,n\}) \setminus \sigma(A), \end{cases} \tag{4.7}$$

$$\sum_{k \in \sigma_i(A)} a'_{ik} = s_i \qquad \text{for} \quad i = 1,\dots,m, \tag{4.8}$$

and

$$\sum_{k \in \sigma^j(A)} a'_{kj} = t_j \qquad \text{for} \quad j = 1,\dots,n. \tag{4.9}$$

Of course, this equivalence is obtained by using the (reversible) change of variables where

$$a'_{ij} = \begin{cases} \ln A'_{ij} & \text{if } (i,j) \in \sigma(A), \\ 0 & \text{if } (i,j) \in (\{1,\dots,m\} \times \{1,\dots,n\}) \setminus \sigma(A), \end{cases} \tag{4.10}$$

$$u_i \equiv \ln U_{ii} \qquad \text{for} \quad i = 1,\dots,m, \tag{4.11}$$

and

$$v_j \equiv \ln V_{jj} \qquad \text{for} \quad j = 1,\dots,n. \tag{4.12}$$

We next observe that the matrix a can be identified with the vector whose coordinates correspond to the nonzero coordinates of the matrix A, listed in lexicographic order. In particular, the problem of finding u, v, and a' satisfying (4.7)–(4.9) can be formulated by (3.1)–(3.2) where a' is also identified with a vector having the same size as a, C is the node-arc incidence matrix of the bipartite graph associated with the matrix A, and $b \equiv (r, s) \in R^m \times R^n = R^{m+n}$. Thus, the results of Section 3 apply. In particular, we will consider the following optimization problem:

PROGRAM II. $\min 2^{-1}\Sigma_{(i,j)\in\sigma(A)}(x_{ij} - a_{ij})^2$

subject to
$$\sum_{j \in \sigma_i(A)} x_{ij} = s_i \qquad \text{for} \quad i = 1,\dots,m, \tag{4.13}$$

$$\sum_{i \in \sigma^j(A)} x_{ij} = t_j \qquad \text{for} \quad j = 1,\dots,n. \tag{4.14}$$

Of course, Program II can be defined explicitly in terms of the original matrix A and the original vectors S and T, but in this case the resulting constraints will not be linear.

We next use the results of Section 3 to study the scaling problem defined by (4.1)–(4.3).

THEOREM 4.1 (Characterization). *Let $A' \in R^{m \times n}$. Then there exist diagonal matrices $U \in R^{m \times m}$ and $V \in R^{n \times n}$ with positive diagonal elements such that (A', U, V) satisfies* (4.1)–(4.3) *if and only if a' defined by* (4.5) *is an optimal solution of Program* II.

Proof. The equivalence follows directly from Theorem 3.1 with C as the arc-node incidence matrix of the bipartite graph associated with A and $b = (r, s) \in R^m \times R^n = R^{m+n}$. The specific arguments make use of the transformation of (4.1)–(4.3) into (4.7)–(4.9), the representation of the latter by (3.1)–(3.2), and the identification of the matrices a and a' with the corresponding vectors. ∎

THEOREM 4.2 (Existence). *The following are equivalent:*

(a1) *there exists a solution to* (4.1)–(4.3),

(a2) *there exists a solution to* (4.7)–(4.9),

(b1) *there exists a nonnegative matrix $A' \in R^{m \times n}$ which satisfies* (4.2)–(4.3) *and has $\sigma(A') = \sigma(A)$,*

(b2) *Program* II *is feasible,*

(c) *Program* II *has an optimal solution,*

(d1) *if $\mu \in R^m$ and $\eta \in R^n$ satisfy $\mu_i = \eta_j$ for all $(i, j) \in \sigma(A)$, then*

$$\prod_{i=1}^{m} (S_i)^{\mu_i} = \prod_{j=1}^{n} (T_j)^{\eta_j}, \tag{4.15}$$

(d2) *if $\mu \in R^m$ and $\eta \in R^n$ satisfy $\mu_i = \eta_j$ for all $(i, j) \in \sigma(A)$, then*

$$\sum_{i=1}^{m} \mu_i s_i = \sum_{j=1}^{n} \eta_j t_j, \tag{4.16}$$

(e1) *if I is a subset of $\{1, \ldots, m\}$ and J is a subset of $\{1, \ldots, n\}$ where $A_{IJ} = 0$ and $A_{I^c J^c} = 0$, then*

$$\prod_{i \in I} S_i = \prod_{j \in J} T_j, \tag{4.17}$$

and

(e2) *if I is a subset of $\{1, \dots, m\}$ and J is a subset of $\{1, \dots, n\}$ where $A_{IJ} = 0$ and $A_{I^c J^c} = 0$, then*

$$\sum_{i \in I} s_i = \sum_{j \in J} t_j. \tag{4.18}$$

Proof. The equivalences (a1) $\Leftrightarrow$ (a2), (b1) $\Leftrightarrow$ (b2), (d1) $\Leftrightarrow$ (d2), and (e1) $\Leftrightarrow$ (e2) are easily verified by taking corresponding natural logarithms as needed. Also, the equivalences (a2) $\Leftrightarrow$ (b2) $\Leftrightarrow$ (c) $\Leftrightarrow$ (d2) follow directly from Theorem 3.2 and the arguments used to establish Theorem 4.1 after observing that (d2) is equivalent to the assertion that if $\mu \in R^m$ and $\xi \in R^n$ satisfy $\mu_i + \xi_j = 0$ for all $(i, j) \in \sigma(A)$, then

$$\sum_{i=1}^{m} \mu_i s_i + \sum_{j=1}^{n} \xi_j t_j = 0. \tag{4.19}$$

Finally, standard results about solvability of the transportation problem (without nonnegativity constraints, show the equivalence of (b2) to (e2), completing our proof. ∎

THEOREM 4.3 (Uniqueness). *There exists at most one matrix $A' \in R^{m \times n}$ for which there exist diagonal matrices $U \in R^{m \times m}$ and $V \in R^{n \times n}$ with positive diagonal elements such that (A', U, V) satisfies* (4.1)–(4.3). *Further, if (A', U^1, V^1) satisfies* (4.1)–(4.3), *then the general solution of* (4.1)–(4.3) *is (A', YU^1, ZV^1) where Y and Z are diagonal matrices having positive diagonal elements and*

$$Y_{ii} Z_{jj} = 1 \qquad \textit{for all} \quad (i, j) \in \sigma(A). \tag{4.20}$$

Proof. Theorem 3.3 and the arguments used in the proof of Theorem 3.1 show the uniqueness of A' and the fact that if (A', U^1, V^1) satisfies (4.1)–(4.3), then the general solution of (4.1)–(4.3) is (A', YU^1, ZV^1), where Y and Z are diagonal matrices having positive diagonal elements such that the vectors $y \in R^m$ and $z \in R^n$ defined elementwise by $y_i = \ln Y_{ii}$ and $z_j = \ln Z_{jj}$ satisfy

$$y_i + z_j = 0 \qquad \text{for all} \quad (i, j) \in \sigma(A).$$

But, the above condition about y and z is clearly equivalent to (4.20), completing our proof. ∎

Consider the bipartite graph $G = (N, E)$ associated with A. We say that two vertices i and j of N are *connected* if there exists a positive integer r and vertices $i_0 = i, i_1, \ldots, i_r = j$ such that for $k = 0, \ldots, r-1$, $\{i_k, i_{k+1}\} \in E$. It is well known that this relation is an equivalence relation and that we get a partition of N into equivalence classes of connected vertices. These equivalence classes are called the *components* of G.

Theorem 4.3 implies that the diagonal elements of the scaling matrices coincide for indices corresponding to elements of the same components of the graph that is associated with A. Hence, a general solution to (4.1)–(4.3) is determined by scalars that correspond to these components. We thereby can obtain a representation of the general solution to (4.1)–(4.3) whenever this system is feasible. To obtain this representation, let Γ be the set of components of G. As we are assuming that A has no zero row or zero column, each $\rho \in \Gamma$ satisfies $\rho \cap \{1, \ldots, m\} \neq \varnothing$ and $\rho \cap \{1, \ldots, n\} \neq \varnothing$. For each $\rho \in \Gamma$, let $Y^\rho \in R^{m\times m}$ be the diagonal matrix with $(Y^\rho)_{ii} = 2$ if $i \in \rho$ and $(Y^\rho)_{ii} = 1$ if $i \in \{1, \ldots, m\} \setminus \rho$, and let $Z^\rho \in R^{n\times n}$ be the diagonal matrix with $(Z^\rho)_{jj} = 2$ if $m + j \in \rho$ and $(Z^\rho)_{jj} = 1$ if $m + j \in \{m + 1, \ldots, m + n\} \setminus \rho$.

COROLLARY 4.4 (Representation). *Suppose* (4.1)–(4.3) *is feasible. Then for some fixed nonnegative matrix $A' \in R^{m\times n}$ and diagonal matrices $U \in R^{m\times m}$ and $V \in R^{n\times n}$ having positive diagonal elements, the general solution of* (4.1)–(4.3) *has the form*

$$\left\{\left(A', U\left[\prod_{\rho\in\Gamma}(Y^\rho)^{\alpha_\rho}\right], \; V\left[\prod_{\rho\in\Gamma}(Z^\rho)^{-\alpha_\rho}\right]\right) : \alpha_\rho \in R \textit{ for each } \rho \in \Gamma\right\}.$$

We note that Lemma 3.4 has an analogue for the problem (4.1)–(4.3), but we do not state it explicitly, because it does not seem to be of interest for the study of the scaling problem.

5. SCALINGS OF MATRICES WITH ROW PRODUCTS EQUAL TO CORRESPONDING COLUMN PRODUCTS

Throughout this section let $A \in R^{n\times n}$ be a given nonnegative matrix having no zero row or zero column. We consider the problem of finding a diagonal matrix $W \in R^{m\times m}$ with positive diagonal element such that

$$A' \equiv WAW^{-1} \in R^{n\times n} \tag{5.1}$$

and

$$\prod_{k \in \sigma_i(A)} A'_{ik} = \prod_{k \in \sigma^i(A)} A'_{ki}, \qquad i = 1, \ldots, n. \tag{5.2}$$

We note that by defining the empty product to be zero, we may allow the matrix A to have zero rows or zero columns. But, under this generalization, solvability of (5.1)–(5.2) implies that a row of A is the zero vector if and only if the corresponding column is. Thus, in such cases we can omit these zero rows and columns.

Consider the matrix $a \in R^{n \times n}$ defined elementwise for $i, j = 1, \ldots, n$ by

$$a_{ij} = \begin{cases} \ln A_{ij} & \text{if } \ (i,j) \in \sigma(A), \\ 0 & \text{if } \ (i,j) \in (\{1, \ldots, n\} \times \{1, \ldots, n\}) \setminus \sigma(A). \end{cases} \tag{5.3}$$

Then (5.1)–(5.2) is equivalent to the problem of finding a matrix $a' \in R^{m \times n}$ and a vector $w \in R^n$ such that

$$a'_{ij} = \begin{cases} w_i + a_{ij} - w_j & \text{if } \ (i,j) \in \sigma(A), \\ 0 & \text{if } \ (i,j) \in (\{1, \ldots, n\} \times \{1, \ldots, n\}) \setminus \sigma(A) \end{cases} \tag{5.4}$$

and

$$\sum_{k \in \sigma_i(A)} a'_{ik} = \sum_{k \in \sigma^i(A)} a'_{ki} \qquad \text{for} \quad i = 1, \ldots, n. \tag{5.5}$$

Of course, this equivalence is obtained by using the (reversible) change of variables where

$$a'_{ij} = \begin{cases} \ln A'_{ij} & \text{if } \ (i,j) \in \sigma(A), \\ 0 & \text{if } \ (i,j) \in (\{1, \ldots, n\} \times \{1, \ldots, n\}) \setminus \sigma(A) \end{cases} \tag{5.6}$$

and

$$w_i \equiv \ln W_{ii} \qquad \text{for} \quad i = 1, \ldots, n. \tag{5.7}$$

As we did in Section 4, the matrix a can be identified with the vector whose coordinates correspond to the nonzero coordinates of the matrix A, listed in lexicographic order. In particular, the problem of finding w and a'

satisfying (5.4)–(5.5) can be formulated as (3.1)–(3.2), where a' is also identified with a corresponding vector, C is the node-arc incidence matrix of the directed graph associated with the matrix A, and $b \equiv 0 \in R^n$. Thus, the results of Section 3 apply. In particular, we will consider the following optimization problem:

PROGRAM III. $\min 2^{-1}\Sigma_{(i,j)\in\sigma(A)}(x_{ij} - a_{ij})^2$

$$\text{subject to} \quad \sum_{k\in\sigma_i(A)} x_{ik} = \sum_{k\in\sigma^i(A)} x_{ki} \qquad \text{for} \quad i = 1,\dots,n. \tag{5.8}$$

Of course, Program III can be defined explicitly in terms of the original matrix A.

We next use the results of Section 3 to study the scaling problem defined by (5.1)–(5.2).

THEOREM 5.1 (Characterization). *Let $A' \in R^{n\times n}$. Then there exists a diagonal matrix $W \in R^{n\times n}$ with positive diagonal elements such that (A', W) satisfies* (5.1)–(5.2) *if and only if a' defined by* (5.4) *is an optimal solution of Program* III.

Proof. The equivalence follows directly from Theorem 3.1 with C as the arc-node incidence matrix of the directed graph associated with A, and $b = 0 \in R^n$. The specific arguments make use of the transformation of (5.1)–(5.2) into (5.4)–(5.5), the representation of the latter by (3.1)–(3.2), and the identification of the matrices a and a' with the corresponding vectors. ∎

THEOREM 5.2 (Existence). *There always exists a solution to* (5.1)–(5.2).

Proof. By applying the transformations discussed above, the proof of the theorem follows directly from the equivalence of parts (a) and (b) in Theorem 3.2 and the observation that the zero vector is always a feasible solution of Program III; hence, this program is always feasible. ∎

THEOREM 5.3 (Uniqueness). *There exists at most one matrix $A' \in R^{m\times n}$ for which there exist diagonal matrices $W \in R^{n\times n}$ with positive diagonal elements such that (A', W) satisfies* (5.1)–(5.2). *Further, if (A', W^1) satisfies* (5.1)–(5.2), *then the general solution of* (5.1)–(5.2) *is (A', ZW^1), where Z is a diagonal matrix with positive diagonal elements and*

$$Z_{ii} = Z_{jj} \qquad \textit{for all} \quad (i,j) = \sigma(A). \tag{5.9}$$

Proof. Theorem 3.3 and the arguments used in the proof of Theorem 3.1 show the uniqueness of A' and the fact that if (A', W^1) satisfies (4.1)–(4.3), then the general solution of (4.1)–(4.3) is (A', ZW^1), and Z is a diagonal matrix having positive diagonal elements such that the vectors $z \in R^n$ defined elementwise by $z_i = \ln Z_{ii}$ satisfy

$$z_i - z_j = 0 \qquad \text{for all} \quad (i, j) = \sigma(A). \tag{5.10}$$

But (5.9) is clearly equivalent to (5.10), completing our proof. ■

Consider the directed graph $G = (N, E)$ associated with A. We say that two vertices i and j of N are *connected* if there exists a positive integer r and vertices $i_0 = i,\ i_1, \dots, i_r = j$ such that for $k = 0, \dots, r-1$, either $(i_k, i_{k+1}) \in E$ or $(i_{k+1}, i_k) \in E$. It is well known that this relation is an equivalence relation and that we get a partition of N into equivalence classes of connected vertices. These equivalence classes are called the *strong components* of G.

Theorem 5.3 implies that the diagonal elements of the scaling matrix coincide for indices corresponding to elements of the same strong component of the graph that is associated with A. Hence, a general solution to (5.1)–(5.2) is determined by scalars that correspond to these components. We thereby can obtain representation of the general solution to (5.1)–(5.2) whenever this system is feasible. To obtain this representation, let Γ be the set of strong components of G. For each $\rho \in \Gamma$, let $Z^\rho \in R^{m \times m}$ be the diagonal matrix with $(Z^\rho)_{ii} = 2$ if $i \in \rho$ and $(Z^\rho)_{ii} = 1$ if $i \in \{1, \dots, n\} \setminus \rho$.

COROLLARY 5.4 (Representation). *For some fixed nonnegative matrix $A' \in R^{n \times n}$ and diagonal matrix $W \in R^{n \times n}$ having positive diagonal elements, the general solution of* (5.1)–(5.2) *has the form*

$$\left\{ \left(A', W \left[\prod_{\rho \in \Gamma} (Z^\rho)^{\alpha_\rho} \right] \right) : \alpha_\rho \in R \textit{ for each } \rho \in \Gamma \right\}.$$

Again, we note that Lemma 3.4 has an analogue for the problem (5.1)–(5.2), but we do not state it explicitly, because it does not seem to be of interest for the study of the scaling problem.

6. ALGORITHMS

The scaling problems considered in Section 4 and in Section 5 can be solved by computing an optimal solution of Program II and Program III, respectively. The corresponding optimal solution, say a', can then be used to obtain the scaled matrix A' by the converse of the transformations given in

(4.10) and (5.3), namely,

$$A'_{ij} = \begin{cases} \exp(a'_{ij}) & \text{if} \quad (i,j) \in \sigma(A), \\ 0 & \text{otherwise}. \end{cases} \tag{6.1}$$

The optimal dual solution of Program II can be used to obtain the scaling matrices U and V by similar exponentiation, and the optimal dual solution of Program III can be used to obtain the corresponding scaling matrix W.

Using the row-action framework of Censor and Lent (1981), one can derive simple iterative algorithms for both Program II and Program III. In particular, the quadratic function $\sum_{(i,j) \in \sigma(A)} 2^{-1}(x_{ij} - a_{ij})^2$ is a Bergman function with zone R^n, as characterized in the above reference. Hence, one can apply the general iterative algorithm of Censor and Lent for equality constrained problems and specialize it for the quadratic objective function and the constraints of Program II and Program III. In particular, the algorithm for Program II is a special case of Algorithm 2.5 from Zenios and Censor (1991). (The latter algorithm solves problems with bounded variables.) Readers are referred to Censor and Lent (1981) for the general algorithm and to Zenios and Censor for specializations and reports of computational experience. Here we only state the algorithms.

ROW-ACTION ALGORITHM for Program II.

Step 0 [initialization]. Set $k \leftarrow 0$. Select $u^0 \in R^m$ and $v^0 \in R^n$. Define $a^0 \in R^{m \times n}$ by

$$a^0_{ij} = \begin{cases} a_{ij} - u^0_i - v^0_j & \text{for} \quad (i,j) \in \sigma(A), \\ 0 & \text{for} \quad (i,j) \notin \sigma(A). \end{cases}$$

Step 1 [iterative step over constraints (4.13)]. For $i = 1, 2, \dots, m$ let

$$\rho_i = \left[\|\sigma_i(A)\|\right]^{-1}\left[s_i - \sum_{j \in \sigma_i(A)} a^k_{ij}\right],$$

$$a^{k+1/2}_{ij} = a^k_{ij} + \rho_i \qquad \text{for} \quad j \in \sigma_i(A),$$

$$u^{k+1}_i = u^k_i - \rho_i.$$

Step 2 [iterative step over constraints (4.14)]. For $j = 1, 2, \dots, n$ let

$$\rho_j = \left[\|\sigma^j(A)\|\right]^{-1}\left[t_j - \sum_{i \in \sigma^j(A)} a^{k+1/2}_{ij}\right],$$

$$a^{k+1}_{ij} = a^{k+1/2}_{ij} + \rho_j \qquad \text{for} \quad i \in \sigma^j(A),$$

$$v^{k+1}_j = v^k_j - \rho_j.$$

Step 3. Replace $k \leftarrow k + 1$ and return to step 1.

ROW-ACTION ALGORITHM for Program III.

Step 0 [initialization]. Set $k \leftarrow 0$. Select $w^0 \in R^n$. Define $a^0 \in R^{n \times n}$ *by*

$$a_{ij}^0 = \begin{cases} a_{ij} - w_i^0 + w_j^0 & \text{for } (i,j) \in \sigma(A), \\ 0 & \text{for } (i,j) \notin \sigma(A). \end{cases}$$

Step 1 [iterative step over constraints (5.8)]. For $i = 1, 2, \ldots, n$ let

$$\rho_i = \left[\|\sigma_i(A)\| + \|\sigma^i(A)\|\right]^{-1}\left[\sum_{p \in \sigma^i(A)} a_{pi}^k - \sum_{p \in \sigma_i(A)} a_{ip}^k\right]$$

$$a_{ij}^{k+1} = a_{ij}^k + \rho_i \quad \text{for } j \in \sigma_i(A),$$

$$a_{ji}^{k+1} = a_{ji}^k - \rho_i \quad \text{for } j \in \sigma^i(A),$$

$$w_i^{k+1} = w_i^k - \rho_i.$$

Step 2. Replace $k \leftarrow k + 1$ and return to step 1.

We conclude by observing that both algorithms are suitable for implementation on parallel architectures. In particular, the algorithm for Program II can be implemented by the simultaneous execution of step 1 for all $i = 1, \ldots, m$, followed by the simultaneous execution of step 2 for all $j = 1, \ldots, n$. The algorithm for Program III can be executed in parallel by iterating simultaneously on indices i and j that have the property $[\sigma_i(A) \cup \sigma^i(A)] \cap [\sigma_j(A) \cup \sigma^j(A)] = \emptyset$. Numerical results reported in Zenios and Censor (1991) with the quadratic optimization algorithm for Program II (with bounded variables) indicate that such parallel implementations are very efficient.

REFERENCES

Avriel, M. 1976. *Nonlinear Programming—Analysis and Methods*, Prentice-Hall, Englewood Cliffs, N.J.

Bacharach, M. 1970. *Biproportional Matrices and Input–Output Change*, Cambridge U.P., Cambridge, U.K.

Brualdi, R. A. 1968. Convex sets of matrices, *Canad. J. Math.* 20:31–50.

Brualdi, S., Parter, S. and Schneider, H. 1966. The diagonal equivalence of a non-negative matrix to a stochastic matrix, *J. Math. Anal. Appl.* 16:31–50.

Censor,Y. and Lent, A. 1981. An iterative row-action method for interval convex programming, *J. Optim. Theory Appl.* 34:321–353.

Censor, Y. and Zenios, S. A. 1991. Interval constrained matrix balancing, *Linear Algebra Appl.* 150:393–421.

Eaves, B. C., Hoffman, A., Rothblum, U. G., and Schneider, H. 1985. Line-sum-symmetric scalings of square nonnegative matrices, *Math. Programming Stud.* 15:124–141.

King, B. B. 1981. What is a SAM? A Layman Guide to Social Accounting Matrices, World Bank Staff Working Paper 463.

Knopp, R. S. 1979. Properties of Kruithof's projection method, *Bell System Tech. J.* 58:517–538.

Kruithof, J. 1937. Telefoonverkeersrekening, *Ingenieur* 52:E15–E25.

Lamond, B. and Stewart, N. F. 1981. Bregman's balancing method, *Transportation Res. Part B* 15:239–248.

Menon, M. V. 1968. Matrix links, an extremization problem and the reduction of a non-negative matrix to one with prescribed row and column sums, *Canad. J. Math.* 20:225–232.

Menon, M. V. and Schneider, H. 1969. The spectrum of a nonlinear operator associated with a matrix, *Linear Algebra Appl.* 2:321–334.

Rothblum, U. G. 1989. Generalized scalings satisfying linear equations, *Linear Algebra Appl.* 114/115:765–784.

Rothblum, U. G. 1992. Linear inequality scaling problems, *SIAM J. Optim.*, to appear.

Rothblum, U. G. and Schneider, H. 1989. More on scalings of matrices having prespecified row sums and column sums, *Linear Algebra Appl.* 114/115:737–764.

Rothblum, U. G., Schneider, H., and Schneider, M. H. 1990b. Scalings of matrices which have prespecified row-maxima and column-maxima, submitted for publication in *SIAM J. on Matrix Analysis.*

Rothblum, U. G., Schneider, H., and Schneider, M. H. 1992. Characterization of max-balanced flows, *Discrete Appl. Math.*, to appear.

Schneider, H. and Schneider, M. H. 1990. Towers and cycle-covers for max-balanced graphs, *Congr. Numer.* 73.

Schneider, H. and Schneider, M. H. 1991. Max-balancing weighted directed graphs, *Math. Oper. Res.*, to appear.

Schneider, M. H. 1990. Max-balanced flows, in *Proceedings of the 1st Waterloo Conference on Integer Programming and Combinatorial Optimization.*

Schneider, M. H. and Zenios, S. A. 1990. A comparative study of algorithms for matrix balancing, *Oper. Res.* 38:439–455.

Sinkhorn, R. 1964. A relationship between arbitrary positive matrices and doubly stochastic matrices, *Ann. Math. Statist.* 35:876–879.

Zenios, S. A. and Censor, Y. 1991. Massively parallel row-action algorithms for some nonlinear transportation problems, *SIAM J. Optim.*, to appear.

Received 22 April 1991; final manuscript accepted 4 October 1991